

## Assignment-1

1. Let  $f(x) = e^x \cos 2x$ .

(a) The definition of the Taylor polynomial is as follows:

$$T_n(x) = \sum_{k=0}^n \frac{(x - x_0)^k}{k!} f^{(k)}(x_0)$$

In this case,  $n = 4$ , and  $x_0 = 0$ .

We start by finding  $f^{(1)}(x)$ ,  $f^{(2)}(x)$ ,  $f^{(3)}(x)$ , and  $f^{(4)}(x)$

$$\begin{aligned} f^{(1)}(x) &= -2e^x \sin(2x) + e^x \cos(2x) = e^x(\cos(2x) - 2\sin(2x)) \\ f^{(2)}(x) &= e^x(-2\sin(2x) - 4\cos(2x)) + e^x(\cos(2x) - 2\sin(2x)) = -e^x(4\sin(2x) + 3\cos(2x)) \\ f^{(3)}(x) &= -e^x(8\cos(2x) - 6\sin(2x)) - e^x(4\sin(2x) + 3\cos(2x)) = e^x(2\sin(2x) - 11\cos(2x)) \\ f^{(4)}(x) &= e^x(4\cos(2x) + 2\sin(2x)) + e^x(2\sin(2x) - 11\cos(2x)) = e^x(24\sin(2x) - 7\cos(2x)) \end{aligned}$$

Evaluating these functions at the center point, we get:

$$\begin{aligned} f(0) &= e^0 \cos(0) = 1 \\ f^{(1)}(0) &= e^0(\cos(0) - 2\sin(0)) = 1 \\ f^{(2)}(0) &= -e^0(4\sin(0) + 3\cos(0)) = -3 \\ f^{(3)}(0) &= e^0(2\sin(0) - 11\cos(0)) = -11 \\ f^{(4)}(0) &= -e^0(24\sin(0) - 7\cos(0)) = -7 \end{aligned}$$

We now have all the pieces we need to determine the polynomial expansion.

$$\begin{aligned} P_n(x) &= \frac{x^0}{0!} f(0) + \frac{x^1}{1!} f^{(1)}(0) + \frac{x^2}{2!} f^{(2)}(0) + \frac{x^3}{3!} f^{(3)}(0) + \frac{x^4}{4!} f^{(4)}(0) \\ &= 1 + x - \frac{3}{2}x^2 - \frac{11}{6}x^3 - \frac{7}{24}x^4 \end{aligned}$$

(b) By definition, the derivative form of the remainder is

$$R_n(x) = \frac{(x - x_0)^{n+1}}{(n+1)!} f^{(n+1)}(c)$$

where  $c$  is close to  $x_0$ .

In this case, we have  $n = 4$ ,  $x_0 = 0$  from the previous question.

$$\begin{aligned}
R_n(x) &= \frac{x^5}{5!} f^{(5)}(c) \\
f^{(5)}(c) &= e^x (48\cos(2x) + 14\sin(2x)) + e^x (24\sin(2x) - 7\cos(2x)) \\
&= e^x (41\cos(2x) + 38\sin(2x)) \Big|_{x=c}
\end{aligned}$$

Therefore,

$$R_4(x) = \frac{x^5}{120} e^c (41\cos(2c) + 38\sin(2c))$$

- (c) We know the remainder,  $R_4(x)$  from part (c). Given a closed interval, we can find the maximum error by finding the maximum remainder in this interval. To do this, we must find its derivative, whose zeroes will show all maxima and minima.

$$f^{(6)}(x) = e^x (-82\sin(2x) + 76\cos(2x)) + e^x (41\cos(2x) + 38\sin(2x)) = e^x (117\cos(2x) - 44\sin(2x))$$

Now find zeroes.

$$e^x (117\cos(2x) - 44\sin(2x)) = 0$$

$$117\cos(2x) - 44\sin(2x) = 0$$

$$\frac{117}{44} = \frac{\sin(2x)}{\cos(2x)}$$

$$x = \frac{1}{2} \tan^{-1}\left(\frac{117}{44}\right) \approx 0.606$$

Our candidate points are then  $\{-\frac{\pi}{4}, 0.606, \frac{\pi}{4}\}$

$$f^{(5)}\left(\frac{\pi}{4}\right) = 38 * e^{\frac{\pi}{4}} \approx 83.3$$

$$f^{(5)}\left(-\frac{\pi}{4}\right) = -38 * e^{\frac{\pi}{4}}$$

$$f^{(5)}(0.606) = e^{0.606} (41\cos(2 * 0.606) + 38\sin(2 * 0.606)) \approx 91.6$$

As  $f^{(5)}(0.606)$  is clearly the max, this is what we use for determining the error bound.

$$\max(R_4(x)) = \frac{\left(\frac{\pi}{4}\right)^5}{5!} * 91.6$$

$$\approx 0.228$$

- (d) Figure 1 shows the actual function compared to its 4<sup>th</sup> degree Taylor expansion over the range  $x \in [-\frac{\pi}{4}, \frac{\pi}{4}]$ . Figure 2 shows the maximum error between the two, which agrees with the upper bound calculated in part (c).

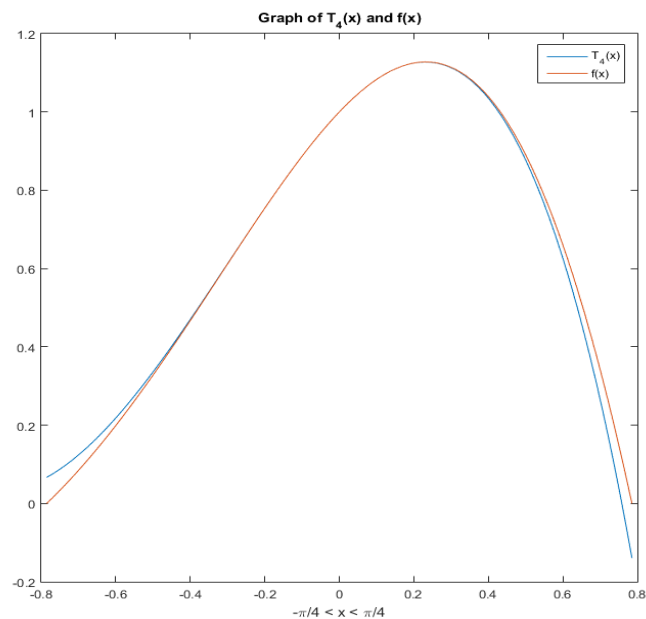


Figure 1: Taylor expansion and initial function

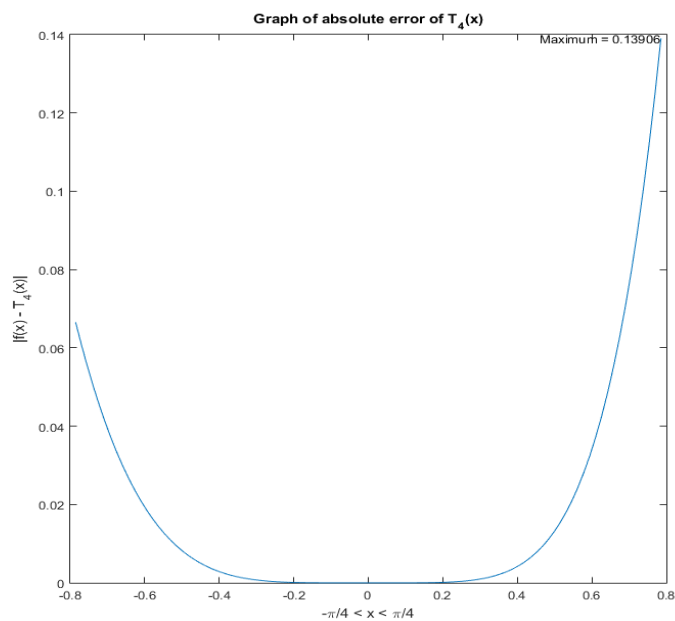


Figure 2: Difference between the Taylor expansion and the initial function

2. (Pencil-and-paper) Convert the binary number  $10.\overline{110}$  to base 10. Give the answer both as a decimal and a fraction if you can.
3. (Pencil-and-paper) Express  $x = 6.7$  as an IEEE single-precision float  $\text{fl}(x)$  using the round-to-nearest rule. Compute the relative error  $d = |x - \text{fl}(x)|/|x|$  exactly as a base-10 number, and show that  $d$  satisfies  $d \leq (1/2)\epsilon_{\text{mach}}$ .
4. (Pencil-and-paper and MATLAB) Holmes 1.5(a).
5. (Pencil-and-paper and MATLAB) Holmes 1.12.
6. (Pencil-and-paper, adapted from Holmes Problem 1.16.) Assume single-precision IEEE arithmetic. Assume that the round-to-nearest rule is used with one modification: if there is a tie then the smaller value is picked (this rule for ties is used to make the problem easier).
  - (a) For what real numbers  $x$  will the computer claim that the inequalities  $1 < x < 2$  hold?
  - (b) For what real numbers  $x$  will the computer claim  $x = 4$ ?
  - (c) Suppose it is stated that there is a floating point number  $x^*$  that is the exact solution of  $x^2 - 2 = 0$ . Why is this not possible? Also, suppose  $x_L^*$  and  $x_R^*$  are the floats to the left and right of  $\sqrt{2}$  respectively. What is the value of  $x_R^* - x_L^*$ ?
7. (a) A problem is ill-conditioned if its solution is highly sensitive to small changes in the input data. True or False?
- (b) Using higher-precision arithmetic will make an ill-conditioned problem better conditioned. True or False?
- (c) If two real numbers are exactly representable as floating-point numbers on a finite-precision machine, then so is their product. True or False?
- (d) Consider the sum

$$S = \frac{1}{x+1} + \frac{1}{x-1}, \quad x \neq 1.$$

For what range of values is it difficult to compute  $S$  accurately in a finite-precision system? How will you rearrange the terms in  $S$  so that the difficulty disappears?

- (e) In a finite-precision system with  $\text{UFL} = 10^{-40}$ , which of the following operations will incur an underflow?
  - i.  $\sqrt{a^2 + b^2}$ , with  $a = 1$ ,  $b = 10^{-25}$ .
  - ii.  $\sqrt{a^2 + b^2}$ , with  $a = b = 10^{-25}$ .
  - iii.  $(a \times b)/(c \times d)$ , with  $a = 10^{-20}$ ,  $b = 10^{-25}$ ,  $c = 10^{-10}$ ,  $d = 10^{-35}$ .